

# Keypoint 検出による植物生育記録の自動化に向けた検討

平原 健太郎<sup>1</sup> 峰野 博史<sup>2</sup>

**概要:** 定期的な生育調査により作物の生育状態を把握することは、栽培と育種の両面において重要である。定期的な生育調査では、植物のストレス状態や成長速度を定量化することで、植物の状態を正確に把握し、過去のデータと比較できる。過去データとの比較は、現在行うべき作物の栽培方法の判断に役立ち、より品質の高い作物の栽培に繋がる。しかし、生育調査は「葉・花のカウント」や「節間の伸長計測」など手作業で行うにはとても手間のかかる作業である。近年の農業従事者の高齢化、新規就農者の減少による労働力不足もあり、従事者は優先度の高い収穫・出荷に追われ、手間のかかる定期的な生育調査を満足に実施できているとは言えない。本研究では、様々な生育調査の中でも「長さ」を測る項目に着目し、Keypoint 検出を用いて自動的に長さ推定を行う技術の確立を目指す。まずは、モデル植物として一般的なトマトの長期多段どり栽培のハウスを対象とし、ハウス内に吊り下げられたロボット撮影機材で栽培ベッド間を撮影した。植物ならではの特徴点・結びつきを定義し、動画から Keypoint 検出によって茎や枝の節を検出することで、茎や枝の長さの自動推定する手法を検証した。さらに、検出した茎を動画内で追跡し、フレーム間で同一個体を結びつかせ個体識別をすることで、より精度の高い長さ推定を可能にした。

## Examination for automation of plant growth record by keypoint detection

KENTARO HIRAHARA<sup>1</sup> HIROSHI MINENO<sup>2</sup>

### 1. はじめに

近年、栽培と育種の両面において、作物の生育状態を的確に把握するための技術開発の重要性が高まっている。作物の生育状態を定期的に調査することで、品質や状態の推定に關係する指標を定量化し、時系列データとして蓄積できる。時系列データとして蓄積した過去の生育状態のデータを比較して栽培評価することで、現在の作物の品質や状態に応じて適切な栽培や環境制御に繋がる。適切な栽培や環境制御で、より品質の高い作物を栽培でき、生育調査は作物の品質を高めるためのとても重要な調査である。

生育調査には、作物の生育状態を把握するため株ごとの節間の伸長計測、果実や花数のカウントなどの様々な調査項目がある。しかし、これらの調査を人手で実施するには手間のかかる作業で、多くの時間と労力を必要とする。さらに、昨今の国内農業分野における高齢化の進行と新規就農者の伸び悩みによって、農業従事者の高齢化と減少も進んでいる。そのため、定期的な生育調査は、品質の高い作物の栽培に重要な作業であるにも関わらず、収穫・出荷といった優先度の高い農作業に追われ、手間のかかる定期的な生育調査は十分実施できているとは言えない。そこで、植物の様々な栽培ステージにおいて、作物の生育状態を定量的に評価する技術が期待される。

本研究では、モデル植物として一般的であるトマトの長期多段どり栽培を対象とし、生育調査業務に含まれる「節間距離」「葉長」といった長さを測る項目について自動で推定する手法を検討する。画像から自動で長さ推定するために Keypoint 検出技術に着目した。Keypoint 検出は、これまで主に人の姿勢検出やハンドトラッキング等に活用されてきた画像解析技術である。トマトのような植物に対して、茎や枝の節といった Keypoint を独自に定義し、植物に対する Keypoint 検出の精度を検証する。さらに、動画からトマトの株および Keypoint を検出し、Object Tracking を用いることで個体認識を行う。Object Tracking で Keypoint の補正を行い、より精度の高い長さ推定を実現する手法を提案する。

以降、第 2 章では関連研究について述べる。第 3 章では提案手法ならびに植物に対する Keypoint の定義、追跡アルゴリズムについて説明し、第 4 章で植物のインスタンス検出精度、Keypoint 推定精度について結果を整理し、第 5 章で本稿をまとめる。

### 2. 関連研究

#### 2.1 画像解析を用いた節間距離測定

トマトの節間距離の推定を行う既存研究として、画像処

<sup>1</sup> 静岡大学大学院総合科学技術研究科  
Graduate School of Integrated Science and Technology, Shizuoka University  
<sup>2</sup> 静岡大学大学院情報学領域  
College of Informatics, Academic Institute, Shizuoka University

理とルールベースの教師あり機械学習を用いた手法[1]がある。この手法では、節検出、節位推定、節間距離推定の3段階の処理によって節間距離を推定している。特に、節検出のために植物画像の全画素に対して「葉」、「茎」、「背景」のラベリングを手作業で行い、決定木ベースのセグメンテーションアルゴリズム Decision Tree based Model(DTSM)[2]を用いて、画素単位での画像分割によって茎領域を抽出し、節の検出を行っている。さらに節位推定では、Bag of Features[3]と呼ばれる一般物体認識手法を用いて誤推定の節位置の削除を行う。このようなルールベースの教師あり機械学習によって、節間距離の推定を行う。しかし、想定している画像は、トマトの苗の茎が明確に映った理想的な画像を対象としており、苗が成長するにつれて葉の量が増加する。そのため、茎や節が成長した葉で隠れてしまった場合、適切に検出できないという課題がある。

## 2.2 Keypoint 検出

Keypoint 検出とは、入力された画像から複数の特徴点を検出し、物体の全体像を Keypoint から推定する技術である。Keypoint 検出を利用した研究として、人間の骨格を推定する OpenPose[4]がある。これは、画像から抽出された人体の関節を Keypoint とし、予め Keypoint 同士を対応付けて繋げる事で骨格の推定を行い、画像から人間の姿勢を推定している。ただし、OpenPose は人間の骨格に着目したものであり、人間以外の対象物には対応していない。

一方、人間以外に着目した Keypoint 検出の研究も行われている。例えば、魚の魚体長の推定に Keypoint 検出を適用した研究[5]がある。この研究では、魚体長の推定のために、口、胸びれ、尾へ Keypoint を独自に定義し、それらを結んだ線分を作成することで魚体長を推定している。つまり、人間以外の物体に対しても新たに Keypoint を定義し、長さの推定を可能とすることを示しており、植物に対しても新たに Keypoint を定義し、長さの推定が可能であるといえる。さらに、動画画像から車両の簡易的な 3D ポリゴンを再生成する CarFusion[6]がある。CarFusion では、車両の角やタイヤに Keypoint を定義することで、街中の道路を走行する車両の検出を実現している。特に Keypoint 検出によって、車長・車幅・車高といった 3次元の情報を 2次元から算出可能としている。つまり、入力画像からは見えない位置にある車体で隠れたタイヤに対しても、Keypoint 検出による特徴点の推定を行えることを示している。

以上のことから、Keypoint 検出技術の特徴は次の2つである。1 つ目にピクセル間の距離を容易に算出可能であること。2 つ目に画像から認識できない隠れた部分を検出できることである。主に人を対象に適用されてきた Keypoint 検出技術を、植物に対して適用することで有効性の検証及び精度評価を行う。対象物に対し適切な Keypoint を定義し、その位置関係を対応付けて学習させることで、隠れた位置

にある Keypoint の検出だけでなく、対象物の長さも適切に推定できると考える。

## 3. 提案手法

### 3.1 想定する条件

本研究では、モデル植物として一般的であるトマトの長期多段どり栽培を対象とする。ハウス内の栽培ベッド間を自走する吊り下げ型ロボットで撮影した動画からデータセットを構築する。苗が成長するにつれて葉量が増えた際に適切に検出できないというルールベースの手法の課題を解決するため、Deep Learning ベースの Keypoint 検出を採用した。本研究では、定期的な生育調査を想定した上で、トマトに対応可能な Keypoint の定義を行い、理想的な画角の画像ではない環境下の画像でも、正確に節間距離や葉長の推定を実現することを目指す。

### 3.2 概要

本研究では、魚体長の推定にも採用された Mask Keypoint R-CNN[7]の後継である Detectron2[8]を用いて、トマトの株の検出および Keypoint の検出を行う。Detectron2 は、次世代の物体検出・セグメンテーションアルゴリズムのフレームワークであり、Facebook AI Research がこれまでに研究開発してきた Detectron と Mask R-CNN のアルゴリズムを継承したフレームワークである。Detectron2 では、先行の Detectron や Mask R-CNN と比べ、学習速度と検出精度の向上を両立している。

本研究では、公開されている Detectron2 の学習済みモデルを利用して、植物用に定義した独自のデータセットからファインチューニングを行い、Keypoint 検出を試みる。学習済みモデルは、COCO[9]の人体骨格検知用のデータセットを用いてトレーニング・評価されたモデルを採用する。学習済みモデルは、人体骨格検知用のモデルではあるが、本研究で想定する植物の茎や枝の節を独自アノテーションした大規模なデータセットを用意することは困難である。そのため、少量の独自アノテーションデータセットで学習できるように、人体骨格検知用の学習済みモデルをファインチューニングした。さらに、検出の難しい隠れた検出箇所を動画から推定するため、追跡アルゴリズムを用いて個体識別を行い、より精度の高い長さ推定を行った。

以下に、本研究の進め方を示す。

#### STEP1. Keypoint 検出の有効性検討

トマトの節間距離の推定を対象とした Keypoint の定義を行う。定義した Keypoint に従って、動画から切り出した画像へアノテーションを行ってデータセットを作成する。データセットを学習用とテスト用に分け、学習用データセットを用いて Detectron2 の学習済みモデ

ルへ Keypoint の学習を行う。

### STEP2. 見えない部分に対する節の検出の検討

データセットの追加・改善や前処理を加えることで、画像には映らない隠れた部分にある節の検出が行えるようモデルの精度向上を図る。ただし、STEP1 同様に、まずは節間距離の推定を対象とする。

### STEP3. 節の検出における精度向上

STEP2 により検出された Keypoint の検出精度向上を図る。Object Tracking により個体を認識し、個体ごとの Keypoint 検出結果から精度の高い節間距離推定を目指す。

### STEP4. 葉長の推定に対応

STEP3 までの結果を通して、葉長推定のための Keypoint を定義し、同様の検出が可能であるか検証する。

### STEP5. 動画へのリアルタイム推定の検討

入力画像の解像度の調整や、モデルの改良等を通じて高速に Keypoint を検出し、距離推定可能か検証する。

## 3.3 節間距離・葉長推定のための Keypoint 定義

Keypoint 検出は主に人間に対して利用され、その定義は人間の各関節に定義されることが多い。その他にも、Deep Learning を用いた Keypoint 検出は様々な分野で応用されており、人間以外の物体にも定義が行われている。しかし、植物に対する Deep Learning を用いた Keypoint 検出を行った事例はこれまでに見当たらない。そのため、本研究では共同研究者との議論を経て、節間距離や葉長推定のための Keypoint を新たに定義することとした。図 1(a)に、定義したトマトの茎と節に対する Keypoint を示す。茎には株の最頂点を示す「Shoot\_apex」、葉の根本である茎の節を「Leaf\_node」として上から 5 点定めた。また、葉には先端を「Leaf\_tip」、葉の中央を「Leaf\_joint」として、「Leaf\_node」

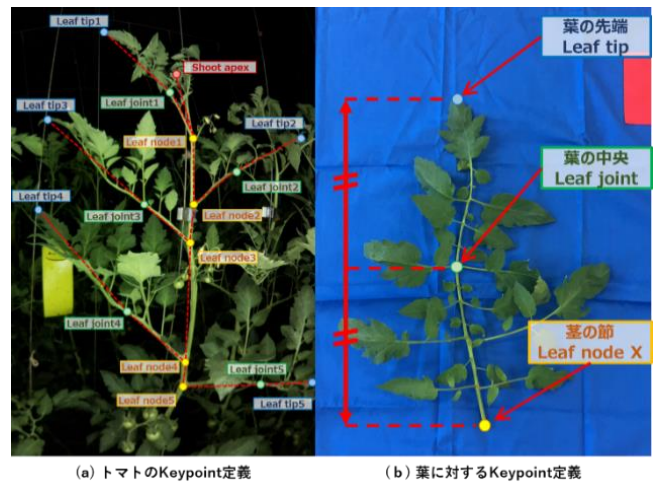


図 1 トマトの茎と節の Keypoint 定義

同様に上から 5 点定めた。

ここで、本研究では図 1(b)に示すよう、主茎に対して分岐する側枝の全体を指すものを「葉」と定義する。一般的な「葉」の定義と異なるが、これは今回想定するトマトの生育調査業務において、側枝の全体を「葉長」として計測しているためである。

## 3.4 検出インスタンスの追跡

動画に対して Detectron2 でフレームごとに Keypoint 検出を行う。フレームごとに検出されたインスタンスを追跡することで個体識別を実現する。個体識別を行うことで Keypoint の誤検出を取り除き、より精度の高い節間距離推定を目指す。

図 2 に提案するトマトにおける Detectron2 の検出を利用した IoU ベースの追跡アルゴリズムを示す。本研究で用いるデータセットは、類似した葉や枝が複数映り込み、色の表現が少ない動画データである。さらに、物体検出に加え

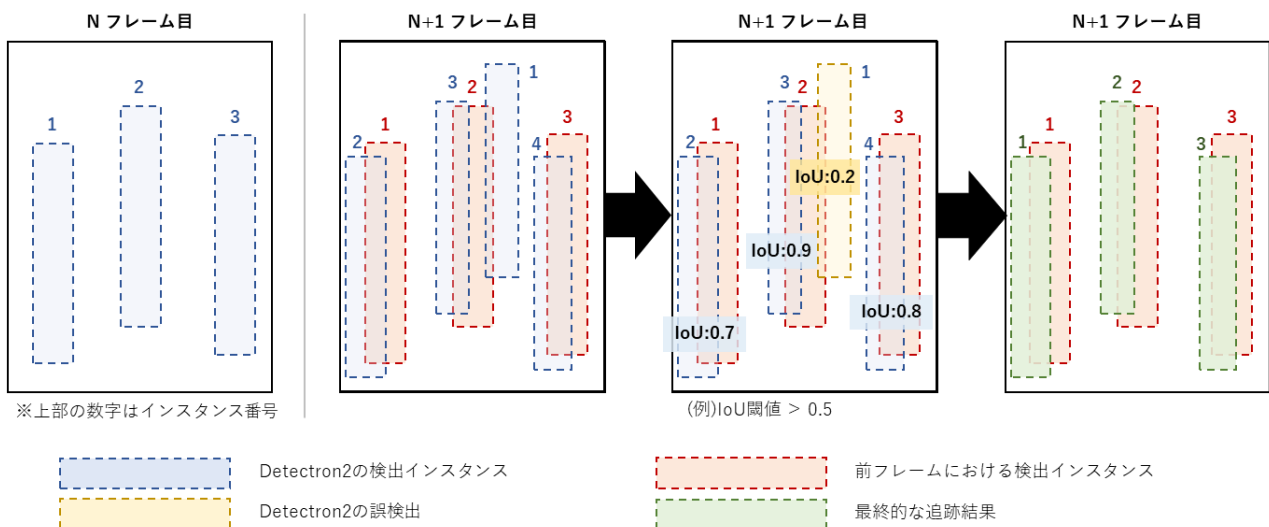


図 2 トマトにおける Detectron2 の検出を利用した IoU ベースの追跡アルゴリズム

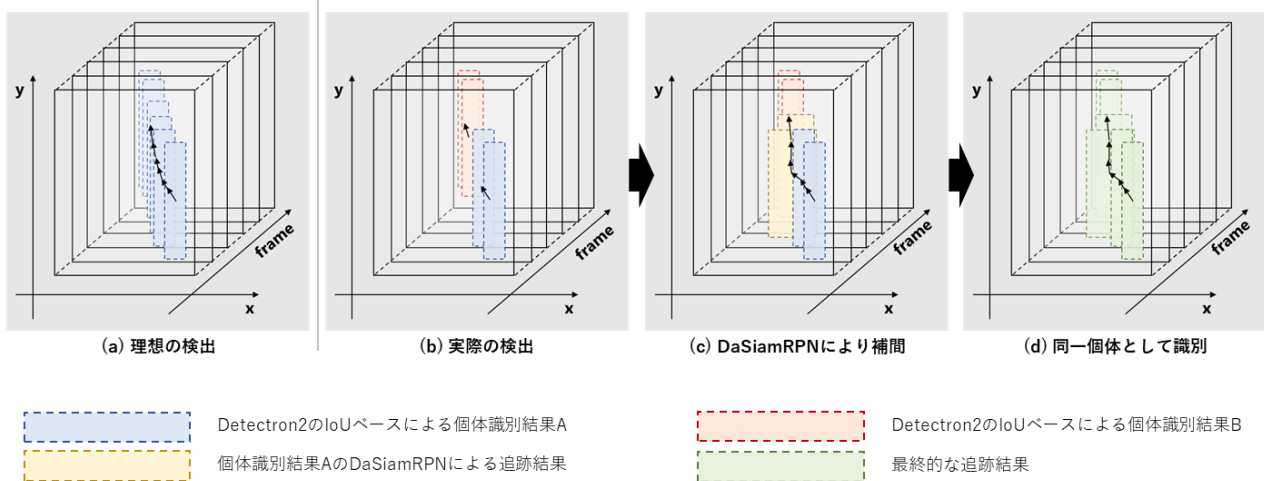


図 3 Detectron2 の検出と DaSiamRPN を用いたインスタンス追跡の流れ

Keypoint の検出も必要となる．そのため，汎用的な FastMOT[10]や FairMOT[11]のような追跡アルゴリズムでは追跡が困難であった．しかし，本研究のデータセットの利点として，ほぼ等間隔で植えられたトマトの株を，一定の高さで横方向に移動し撮影した動画データであることが挙げられる．以上のような，データセットの利点・欠点から IoU Based Tracking[12]をベースにしたアルゴリズムを利用して，Detectron2 の検出インスタンスを追跡することとした．しかし，一連の動画フレームにおいて，同一個体にも関わらず Detectron2 によって検出できないフレームが存在する．このような場合，前述の IoU ベースの追跡アルゴリズムでは追跡が途切れ，同一個体が別々のインスタンスとして識別されてしまう．このような状態を避けるために，高精度ながら高速に追跡可能な DaSiamRPN[13]を使用して，Detectron2 によって検出できないインスタンスを前フレームから検出する．

図 3 に Detectron2 の検出と DaSiamRPN を用いたインスタンスの追跡の流れを示す．Keypoint の検出は，DaSiamRPN を用いて検出したフレームのインスタンスには適用せず，Detectron2 でのみ行った．

## 4. 評価実験

本稿では，前述の STEP3 までの進捗状況について報告する．Keypoint 定義に従って，節間距離の推定に必要なアノテーションデータのみを用いて実験を行う．つまり，本稿で用いるアノテーションデータは，葉に定義される Keypoint を除く，主茎部分に定義する 6 点の Keypoint のみを選択する．

### 4.1 データセット

表 1 にデータセットの詳細を示す．動画は山梨県のハウスで栽培されるトマトの長期多段どり栽培のベッドを対象

に撮影した．撮影にはレールに吊り下げられるロボット撮影機材を利用して，一定の高さで横方向に移動し栽培ベッドの様子を取得した．栽培ベッドの全長は約 120m である．レンズはより広くハウス内のトマトが映るよう魚眼レンズで撮影し，手前の株のみが映るよう夜間に撮影することで背景が黒くなるようにした．図 4 に取得した動画データの

表 1 データセット詳細

撮影	場所	山梨県北州市 有限会社アグリマインド	
	対象	トマト（長期多段どり栽培）	
	年月	2020年8月, 2021年5-7月（計10日間, 夜間に撮影）	
	レンズ画角 本数	165度（魚眼） 22本	
動画	ファイル形式	mp4形式	
	解像度	幅960px, 高さ1280px	
	平均再生時間	約2分	
学習	フレームレート	30fps	
	動画からの切り出し	約2m間隔（50フレーム間隔）	
	アノテーションツール	COCO-Annotator	
	アノテーション	Segmentation + Keypoint	
評価	アノテーション枚数	学習用 420枚 検証用 53枚 評価用 53枚	計526枚
	動画からの切り出し	3フレーム間隔（10fps）	
	アノテーション枚数	Keypointのみ 1134枚	
	動画換算	1分53秒	

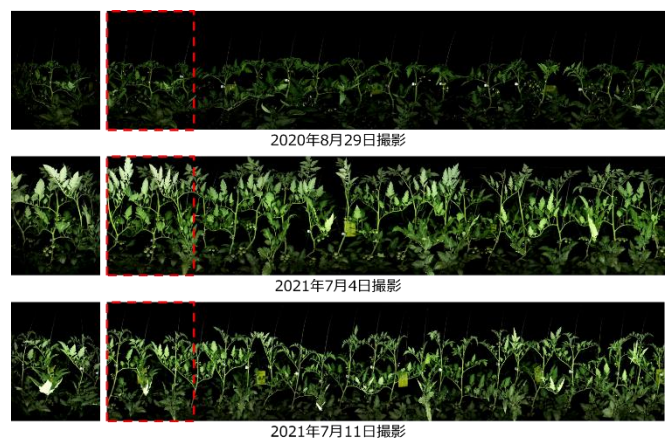


図 4 取得した動画データのパンorama画像



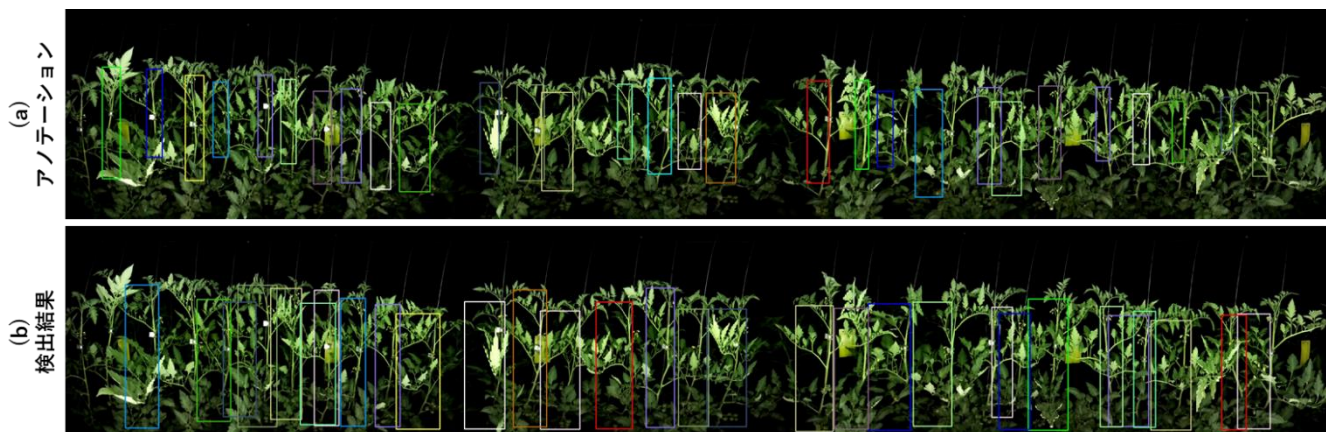


図 5 アノテーション及び検出結果のインスタンスを描画したパノラマ画像

一例としてパノラマ化した画像を示す。

また、Detectron2 学習に用いるアノテーション済み画像データセットを用意した。動画データから約 2m 間隔 (50 フレーム間隔) で画像を切り出し、COCO-Annotator[14]を用いてセグメンテーション及び Keypoint のアノテーションを行った。Detectron2 学習のためのアノテーション済み画像は計 526 枚を用意した。学習には、訓練データ、検証データ、評価データごとに 8 : 1 : 1 に分割して利用した。

最後に、評価用のデータセットについて説明する。この評価用データセットを用いて、推定された節間距離の評価を行う。評価用データセットは撮影した動画データに対して、フレームごとに Keypoint のみアノテーションを行ったものである。アノテーションの手間を考慮し、元動画のフレームレートを 30fps から 10fps に下げて、動画全フレームの 3 分の 1 フレームに対して Keypoint のアノテーションを行った。節間距離評価用のアノテーション済みデータは計 1134 枚、動画にすると 1 分 53 秒になるものを用意した。

#### 4.2 Detectron2 の学習

茎の検出・節の検出を Detectron2 により行う。Detectron2 は計 526 枚のアノテーション済み画像データセットを用いてファインチューニングを行った。使用する Keypoint の学習済みモデルには、Model Zoo[15]にて公開されている「X\_101\_32x8d\_FPN\_x3」を使用した。「X\_101\_32x8d\_FPN\_x3」の学習済みモデルは、ResNet[16]のアーキテクチャを改良したモデルである ResNeXt[17]と呼ばれるモデルを採用している。ResNeXt は cardinality と呼ばれる要素を追加することにより、複雑さを維持しつつ、ResNet のモデルよりも高精度な分類を可能とする。バッチサイズ 128, 学習率 0.0004 とし、10000[iter]のファインチューニングを行った。また、推論の信頼度を 0.80 の閾値に設定した。

#### 4.3 追跡による動画から茎の個体識別

提案手法で示した追跡アルゴリズムによるトマトの茎の個体識別について、動画の長さによる正解データとの検

表 2 それぞれのデータセットにおける検出インスタンス数

分割なし(1分53秒)			5分割 (約22.6秒/1動画)			10分割 (約11.3秒/1動画)		
正解値	検出数	誤差	正解値	検出数	誤差	正解値	検出数	誤差
139	147	+8	31	32	+1	14	17	+3
			28	29	+1	19	16	-3
			29	28	-1	16	14	-2
			29	30	+1	15	16	+1
			31	30	-1	15	16	+1
						15	14	-1
						17	15	-2
						15	16	+1
						17	15	-2
						16	16	0

出インスタンス数の比較について行った。評価は表 1 に示した節間距離推定のための評価用データセットを用いて評価を行う。動画のフレームを 5 分割 (約 22.6 秒 / 1 動画)、10 分割 (約 11.3 秒 / 1 動画) したデータセットを用意し、3 種類のデータセットにおける検出インスタンス数を調査した。

表 2 に、それぞれのデータセットにおける検出インスタンス数の結果を示す。表 2 の結果より、追跡アルゴリズムは完璧とはいかないまでも最大誤差 8 つまでの範囲でインスタンスの追跡に成功したことが分かった。表 1 から、Detectron2 の学習に使用したアノテーション済み画像は 526 枚と少なく、植物という形の定まらない検出対象でありながらも、概ね良好に動画に映る茎の個体を識別できていることが確認できる。さらに、5 分割したデータセットでは、インスタンス数の誤差が ±1 の範囲にあり、高精度な個体識別が行えていることが示された。図 5 に例として、5 分割したデータセットのうち 1 つから、茎のアノテーション位置と Detectron2 によって検出した位置を描画する動画をパノラマ画像にしたものを示す。図 5 は動画から作成したパノラマ画像であることから、描画するバウンディングボックスの X 座標がやや左右にずれてしまう描画上の不具合があるが、おおよその正解位置と検出位置の位置関係

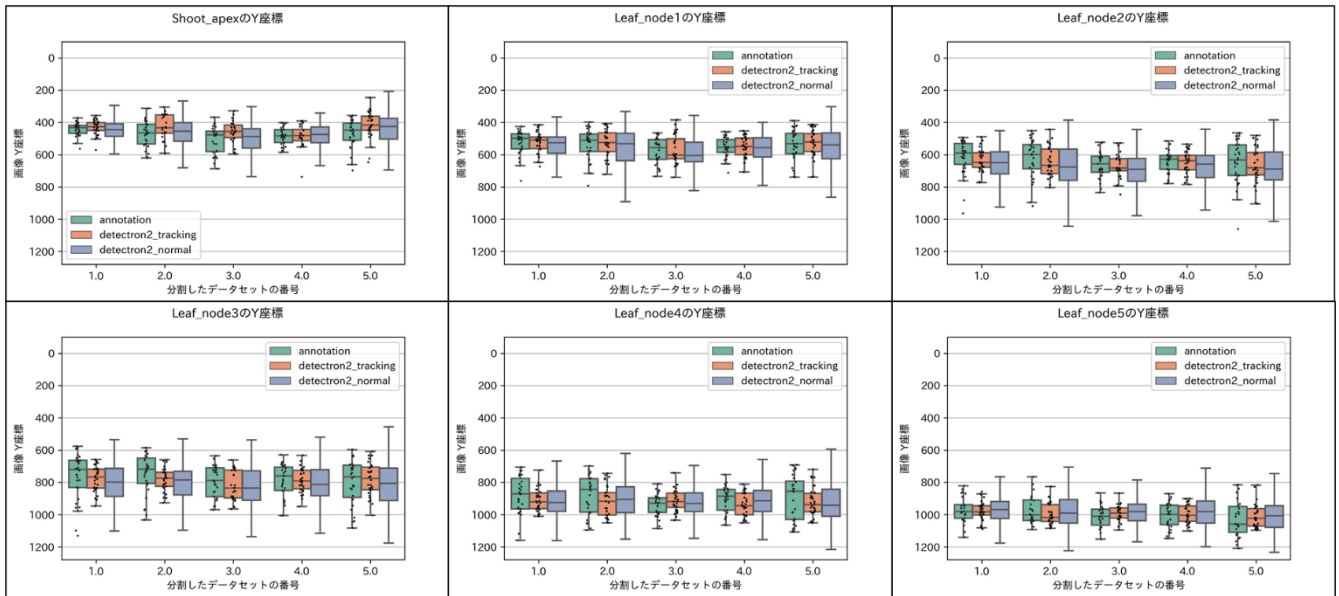


図 6 検出した Keypoint それぞれの Y 座標の分布

について考察する．図 5 より，動画端にある茎個体の未検出や茎が密集している箇所に関して重複して検出していることがわかる．動画端のインスタンスに関しては，動画データを分割していることから，本来中心に映るべき茎が端に写ってしまっていることが未検出となる原因である．そのため，実運用では茎を中心を通るように撮影することで正しく検出することができると考える．茎が密集している箇所に関しては，追跡アルゴリズムの改良や，Detectron2 の学習に利用するデータセットの拡充が必要となるのではないかと考える．

#### 4.4 インスタンス追跡による効果

学習した Detectron2 を利用して，提案手法のインスタンス追跡を調査した．データセットには，前節にて最もインスタンス検出性能の高かった評価用動画を 5 等分割したものをを用いた．比較対象は「アノテーションの正解値 (annotation)」、「Detectron2 の出力結果 (detectron2\_normal)」、「Detectron2 の出力結果を利用してインスタンス追跡を実施した結果 (detectron2\_tracking)」の 3 パターンである．それぞれのパターンにおいて，5 つの動画から検出した節 6 種類 (Shoot\_apex 及び Leaf\_node\_1-5) の Y 座標の分布を分析した．

図 6 に，検出した Keypoint それぞれの Y 座標の分布を示す．図 6 の結果より，Detectron2 の検出結果を利用してインスタンス追跡を行った場合，元の Detectron2 の出力結果と比べ Y 座標の分布のばらつきが小さくなり，正解値の Y 座標の分布に類似していることがわかる．このことから，インスタンス追跡によって，元のデータから Y 座標が外れ値となる Keypoint が取り除かれ，より正解の節と近い検出 Keypoint が選別されていると考える．また，Keypoint の分布のばらつきが小さいため，重複して同じ節を検出するよ

うな，誤検出した Keypoint を取り除くことに成功しているといえる．

続いて，図 7 にインスタンス追跡によって検出した同ラベルの Keypoint を動画のフレーム間で結びつけた例を示す．図 7 は上から Shoot\_apex, Leaf\_node\_1, Leaf\_node\_2, Leaf\_node\_3, Leaf\_node\_4, Leaf\_node\_5, の Keypoint を表し，それぞれの Keypoint で検出した同一インスタンス内の Y 座標の中央値に線を引いたものである．図 7 より，同一インスタンスであっても Keypoint 検出にばらつきが見られることがわかる．特に，Leaf\_node\_4 の節でありながらも Leaf\_node\_5 が同じ位置に検出されるなど，重複して同じ節を検出しているフレームが多々あることが見受けられる．しかし，図 7 中に描画した Keypoint ごとの中央値の線に着目すると，正しく節の位置を推定していることがわかる．つまり，重複して誤検出された Keypoint を外れ値として除去することに成功していると考えられる．したがって，追跡アルゴリズムで個体を識別し，個体ごと Keypoint の分布を統計的に処理することで，重複して誤検出した Keypoint を除去し，より精度の高い節位置の検出を実現できると考える．

#### 5. おわりに

本研究では，トマトの長期多段どり栽培において，生育調査業務に含まれる「節間距離」「葉長」といった長さを測る項目において，Keypoint 検出という観点から自動化を行う手法を検討した．提案手法では，栽培の様子を自動走行する車両から一定の高さで横方向に撮影し，Keypoint 検出技術によって「節間距離」「葉長」の検出を行った．さらに，撮影する角度によって検出結果が異なることを利用し，動画からの個体識別によって誤検出を除去することを提案した．本稿では，まず「節間距離」推定のため，節の Keypoint

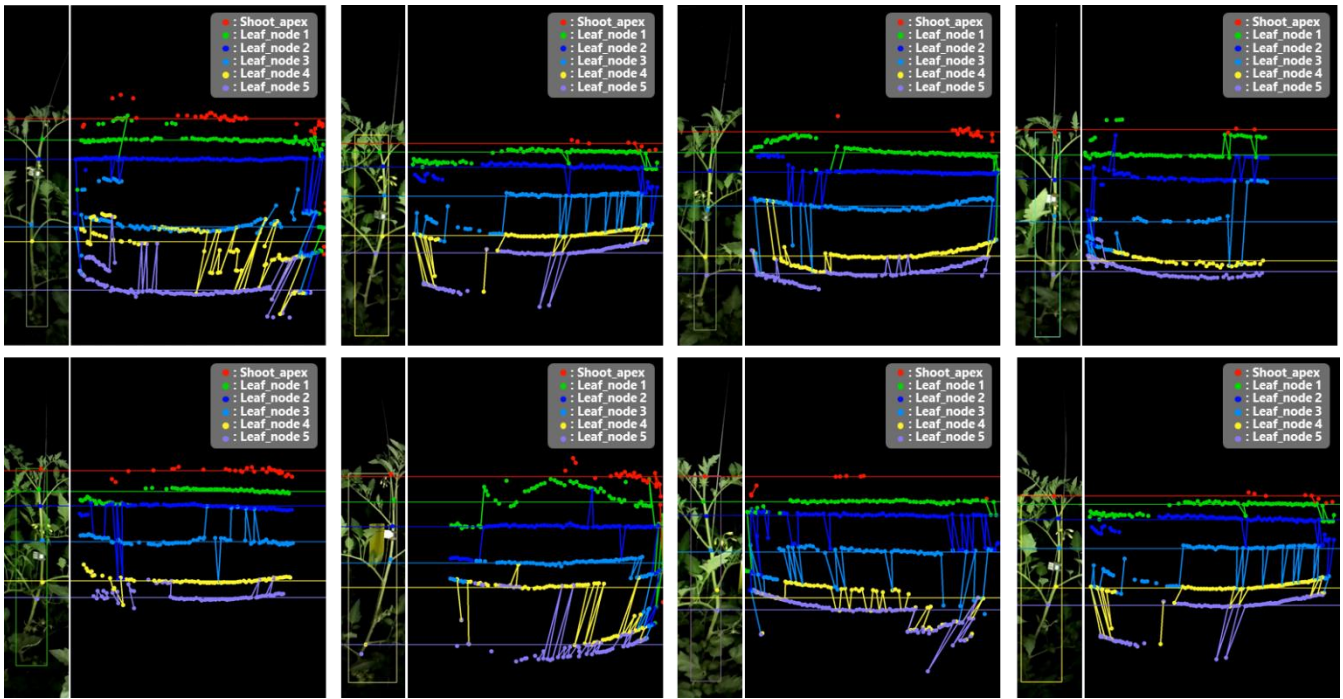


図 7 検出インスタンスにおける Keypoint の軌跡

検出に焦点を絞って検証を進めた。実験の結果、トマトの栽培ペットを横方向に移動して撮影した動画から、追跡アルゴリズムを使うことによって、Detectron2 により茎部分の検出、節の検出を高精度に行えることを確認できた。追跡アルゴリズムは、重複して誤検出した Keypoint を除去し、より正解に近い節位置の推定を可能にした。

今後の方針として、検出インスタンスの重複を防ぐために追跡アルゴリズムの改良を行う。現在は IoU ベースの追跡アルゴリズムを利用しているが、IoU ベースの追跡では茎が密集している箇所重複検出のリスクが高まる。密集区域での個体識別を実現できるように、追跡アルゴリズムの改良等を行い検証をすすめる。さらに、少量データセットで高精度な検出を行えるよう画像増幅手法の検討や、実運用に対応できるように検出精度や検出速度の向上の検討を行っていく。

## 謝辞

本研究の一部は、JST 創発的研究支援事業(JPMJFR201B)の支援を受けたものである。また、データセットの提供及び農学知見を支援いただいた株式会社 PLANT DATA の加納様、海野様、北川様、NTT 社会情報研究所の工藤様、永井様に深い感謝の意を表する。

## 参考文献

[1] 山本恭輔, "画像解析と機械学習によるトマトの自動生育診断および高速フェノタイピングに関する研究", 東京大学大学院農学生命科学研究科博士論文 (2015).  
 [2] Guo, W., Rage, U. K., and Ninomiya, S., "Illumination invariant segmentation of vegetation for time series wheat images based on

decision tree model", Computers and Electronics in Agriculture, Vol. 96, pp. 58-66 (2013).  
 [3] Csurka, G., Dance, C. R., Fan, L., Willamowski, J., Bray, C., "Visual categorization with bags of keypoints", Proceedings of the ECCV International Workshop on Statistical Learning in Computer Vision, pp. 59-74 (2004).  
 [4] Zhe, C., Gines H., Tomas S., Shih-En, W., Yaser, S., "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 1, pp. 172-186 (2019).  
 [5] 長谷川達人, 田中基貴, "水産資源管理に向けた Mask Keypoint R-CNN による魚体長認識", 情処研報 Vol.2021-CDS-32, No.12 (2021).  
 [6] N Dinesh Reddy, Minh Vo, Srinivasa G. Narasimhan, "CarFusion: Combining Point Tracking and Part Detection for Dynamic 3D", 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1906-1915 (2018).  
 [7] He, K., Gkioxari, G., Doll'ar, P., Girshick, R, "Mask R-CNN", 2017 IEEE International Conference on Computer Vision (ICCV), pp. 2980-2988 (2017).  
 [8] FaceBookAI, "Detectron2", (2021).  
 <<https://github.com/facebookresearch/detectron2/>> (参照 2022-05-21)  
 [9] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Doll'ar, C. L. Zitnick, "Microsoft COCO: Common Objects in Context", European Conference on Computer Vision (ECCV), pp. 740-755 (2014)  
 [10] Yukai, Y., "FastMOT", (2021)  
 <<https://github.com/GeekAlexis/FastMOT>> (参照 2022-05-21)  
 [11] Yifu, Z., Chunyu, W., Xingang, W., Wenjun, Z., Wenyu, L., "FairMOT: On the Fairness of Detection and Re-Identification in Multiple Object Tracking", International Journal of Computer Vision (IJCV) pp. 1-19 (2021)  
 [12] Erik, B., Tobias, S., Thomas, S., "Extending IOU Based Multi-Object Tracking by Visual Information", IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) pp. 1-6 (2018)  
 [13] Zheng, Z., Qiang, W., Bo L., Wei W., Junjie Y., Weiming H.,

- "Distractor-aware Siamese Networks for Visual Object Tracking",  
ECCV2018, pp. 101-117 (2018).
- [14] Justin Brooks, "COCO Annotator", (2019)  
<<https://github.com/jsbroks/coco-annotator/>> (参照 2022-05-21)
- [15] Jing Yu Koh. "Model Zoo"  
<<https://modelzoo.co/>> (参照 2022-05-21)
- [16] Kaiming, H., Xiangyu, Z., Shaoqing, R., Jian, S., "Deep Residual Learning for Image Recognition", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778 (2016)
- [17] Saining, X., Ross, G., Piotr, D., Zhuowen, T., Kaiming, H.,  
"Aggregated Residual Transformations for Deep Neural Networks",  
IEEE Conference on Computer Vision and Pattern Recognition  
(CVPR), pp. 1492-1500 (2017).